Project Abstract

Integrated International Microdata Series
SES-0433654
Steven Ruggles
University of Minnesota

A vast body of raw census microdata covering much of the world over the past four decades survives in machine-readable form. The bulk of these data, however, remains inaccessible to researchers. This proposal seeks funding to create an integrated global database of over 150 censuses from at least 44 countries. The International Integrated Public Use Microdata Series (IPUMS-International) will be the world's largest public-use population database, with multiple samples from each country enabling analyses across time and space. These microdata and accompanying documentation will be freely available for scholarly and educational research through a web-based data dissemination system.

The project leverages NSF investment in a major social science infrastructure project now nearing completion, "International Integrated Microdata Access System" (SBR9908380). That project covered many of the costs of finding and preserving microdata and documentation, negotiating dissemination agreements, developing data cleaning and sampling procedures, creating data conversion and dissemination software, and establishing design protocols for data and documentation. As a result, creation of the new database will be highly cost effective.

Census microdata represent an extraordinary untapped resource for research and education in human and social dynamics. With over five hundred million records spanning four decades, the new database will offer far broader chronological scope and greater sample densities than any alternative data source. For most countries, censuses are the most representative source of population data available. The new database will allow investigators to analyze global change during a period of unprecedented economic, demographic, and political upheaval.

Progress Report

We began work on the project in January 2005. We have two main goals for the first year of the project. First, we are working to preserve and democratize access to as much census microdata as possible; second, we are developing a comprehensive design for creating the database.

Our data preservation efforts have been remarkable successful. We have successfully executed dissemination agreements with 44 countries, and have acquired both the data (over one billion person records) and dissemination licenses which allow us to redistribute it. This represents the minimum number of countries we promised in the proposal, but our work on preservation continues. By the end of the project, we hope to have ensured the survival and usability of many additional microdata collections.

The development of a streamlined work plan is also proceeding well. The new plan represents a significant restructuring of the work processes we employed in the first phase of this research.

Our new approach improves scalability, speeds processing, and reduces the cost of database maintenance. Highlights of the work process include 1) full implementation of XML metadata standards, 2) elimination of metadata redundancies, 3) all new java-based data conversion software, with increased analyst control, 4) improvement of the data access tools. We will present a final version of the design at the annual meeting of our Advisory Board in April 2006.

Broader Impact

This project will reduce barriers to international research and education by preserving datasets and making them freely available, converting them into a uniform format, providing comprehensive documentation, and implementing web-based tools for disseminating the microdata and documentation. The database will provide fundamental infrastructure for a broad range of fields in the social and behavioral sciences, including economics, geography, sociology, population studies, and environmental studies. Researchers in most countries do not presently have access even to their own national census microdata; IPUMS-International will democratize access to this vital scientific resource, creating unprecedented opportunities for global-scale research.

Most census data have traditionally been available only in aggregated tabular form. Census *microdata* provide information about individual persons, families, and households, and they allow users to interrelate any desired set of population and housing characteristics. The flexibility offered by microdata is essential for comparative research on social dynamics because the aggregate tabulations produced by national statistical offices are usually not comparable across time or between countries. In the few countries where census microdata covering multiple census years have been easily available to researchers, these data are the most widely used source for the study of large-scale economic and demographic transformations. Making integrated census microdata available for almost half of the world's population will allow researchers to describe the transformation of the world in far richer detail than previously possible. Even more important, these data will provide unprecedented opportunities to investigate the agents of change and assess their implications for human society.

Census microdata are an essential resource for studying large-scale transformational changes such as economic development, urbanization, fertility transition, large-scale migration, population aging, mass education, democratization, and growing international trade and capital flows. The availability of multiple censuses for each country lends historical depth, revealing the trajectories of change hidden in snapshots from the recent past. These data allow detailed study of the relationships of social and economic change to variations in climate, geography, and environment. They are also uniquely suited to assessing the human consequences of social, economic, and demographic transformations in such diverse areas as family structure, economic inequality, cultural diversity, and assimilation.

Project Website

http://ipums.org/international